NOTES

ON A NEW METHOD OF COMPUTING FREQUENCY PERCENTAGES

By Robert E. Moritz, *University of Washington*

Suppose there be given the frequencies

$$f_1, f_2, f_3, \ldots, f_n,$$

whose sum is

$$F = f_1 + f_2 + f_3 + \ldots + f_n,$$

and it is required to compute the frequency percentages

$$p_k = 100 f_k / F, \quad k = 1, 2, 3, \ldots, n.$$

The method commonly employed is to compute the ratios f_k/F by means of a slide-rule, or, if a slide-rule is not available, to compute the ratio $100/F$ and then multiply, logarithmically or otherwise, each of the numbers f_k by this common factor.

Some time ago, while spending a vacation period at an island camp, the writer unexpectedly interested himself in a statistical inquiry which involved the computation of several hundred tables of frequency percentages. Being deprived of the usual aids in computation, he hit upon the following method of approximation, which, as far as simplicity and rapidity is concerned, makes the usual aids almost, if not entirely, superfluous. Moreover, the writer is now able to show that the results obtained by this method are sufficiently accurate for all practical purposes—the method admitting, in fact, of extension so as to secure any desired degree of accuracy.

The method referred to consists of two steps.

(1) The frequencies f_k are replaced by proportional numbers f'_k whose sum F' approximates 100.

(2) The difference $D = 100 - F'$ is distributed over the numbers f'_k in accordance with the formula $\Delta f'_k = f'_k \cdot D/100$, where $\Delta f'_k$ is that portion of the difference D which falls to f'_k.

It is obvious that when $F' = 100$, there remains no difference to be distributed, and the numbers f'_k are the true frequency percentages. If, however, F' differs from 100, as is usually the case, the numbers f'_k, augmented as directed in (2), will differ slightly from the true frequency percentages, their true values being $f'_k + f'_k \cdot D/F'$, while our formula gives $f'_k + f'_k \cdot D/100$.

We shall illustrate the method by some examples before we consider the magnitude of the errors involved.

In the tables that follow, f_k represents the given frequencies, f'_k the derived frequencies obtained as directed in (1), $\Delta f'_k = f'_k \cdot D/100$, the portions of the difference $D = 100 - F'$ to be added to f'_k and $p'_k = f'_k + \Delta f'_k$, the approximate frequency percentages.

TABLE I

k	f'_k	f'_k	$\Delta f'_k$	p'_k
1	4	6.	+0.06	6.06
2	15	22.5	0.225	22.725
3	23	34.5	0.345	34.845
4	17	25.5	0.255	25.755
5	7	10.5	0.105	10.605
	$F=66$	$F'=99.0$		99.99 (check)

(1) $f'_k = f_k + \frac{1}{2} f_k$. (2) $100 - F' = 1$. $\Delta f'_k = 1\%$ of f'_k.

TABLE II

k	f_k	f'_k	$\Delta f'_k$	p'_k
1	5	6.	−0.05	5.95
2	10	12.	0.10	11.90
3	13	15.6	0.13	15.47
4	24	28.8	0.23	28.57
5	16	19.2	0.15	19.05
6	9	10.8	0.09	10.71
7	7	8.4	0.07	8.33
	$F=84$	$F'=100.8$		99.98 (check)

The results have been rounded off to two decimal places.

(1) $f'_k = f_k + \frac{1}{5} f_k$. $100 - F' = -0.8$. $\Delta f'_k = -\frac{8}{10}\%$ of f'_k.

TABLE III

k	f_k	f'_k	$\Delta f'_k$	p'_k
1	9	5.4	−0.14	5.26
2	24	14.4	0.36	14.04
3	30	18.0	0.45	17.55
4	40	24.0	0.60	23.40
5	31	18.6	0.46	18.14
6	27	16.2	0.40	15.80
7	10	6.0	0.15	5.85
	$F=171$	$F'=102.6$		100.04 (check)

(1) $f'_k = \frac{6}{10} f_k$. (2) $100 - F' = -2.6$, $\Delta f'_k = -\frac{1}{4}$ of 10% of f'_k (nearly).

TABLE IV

k	f_k	f'_k	$\Delta f'_k$	p'_k
1	18	0.72	+0.005	0.725
2	42	1.68	0.012	1.692
3	85	3.40	0.024	3.424
4	101	4.04	0.028	4.068
5	150	6.00	0.041	6.041
6	202	8.08	0.055	8.135
7	300	12.00	0.082	12.082
8	450	18.00	0.122	18.122
9	373	14.92	0.101	15.021
10	252	10.08	0.069	10.149
11	178	7.12	0.048	7.168
12	120	4.80	0.033	4.833
13	90	3.60	0.025	3.625
14	59	2.36	0.016	2.376
15	50	2.00	0.014	2.014
16	13	0.52	0.004	0.524
	$F=2483$	$F'=99.32$		99.999 (check)

(1) $f'_k=\frac{4}{100}f_k$. (2) $100-F'=0.68=0.7$ nearly. $\Delta f'_k=\frac{7}{10}$ of 1% of f'_k (nearly).

The values of p'_k are correct to three decimal places For many purposes the values f'_k give sufficiently close approximations of the values p_k.

When the method is once clearly understood, the corrections $\Delta f'_k$ need not be tabulated but may be immediately applied to the f'_k's and the results entered in the column p'_k.

Let us now consider the magnitude of the errors introduced into the values of p'_k because of (2). Let p_k represent the true values of the frequency percentages, and let

$$p_k=f'_k+c\cdot f'_k,$$

where c is the fraction of f'_k that must be added to f'_k to obtain p_k. On summing all the values of p_k we have

$$\Sigma p_k=\Sigma(f'_k+c\cdot f'_k)=\Sigma f'_k+c\cdot\Sigma f'_k,$$

from which, since $\Sigma p_k=100$, and $\Sigma f'_k=F'$,

$$100=F'+c\cdot F'$$

that is

$$c=(100-F')/F'.$$

The true frequency percentages are therefore given by the formula

$$p_k=f'_k+f'_k(100-F')/F',$$

while

$$p'_k=f'_k+f'_k(100-F')/100.$$

We may now calculate the error in p'_k, namely,

$$p_k - p'_k = f'_k(100-F')\left(\frac{1}{F'} - \frac{1}{100}\right) = f'_k \frac{(100-F')^2}{100F'};$$

also, the sum of all the errors

$$100 - \Sigma p'_k = \frac{(100-F')^2}{100F'} \Sigma f'_k = \frac{(100-F')^2}{100},$$

and the relative error

$$R = \frac{100 - \Sigma p'_k}{100} = \left[\frac{100-F'}{100}\right]^2.$$

The last formula shows that in the case of the four examples given, the relative errors do not exceed 0.0001, 0.00007, 0.0007, and 0.00005 respectively. F' may therefore vary from 100 within much wider limits than in the case of the examples cited and yet yield values of p_k sufficiently accurate for all practical purposes. If F' is anywhere between 95 and 105, the relative error will be less than $\frac{1}{4}$ of 1 per cent. Moreover, it is easy to determine beforehand the limits within which F' must be chosen so as to give any preassigned accuracy to the values of p_k. For example, if the relative error is to be less than $\frac{1}{10}$ of 1 per cent, we have $\left[\frac{100-F'}{100}\right]^2 < 0.001$, from which $96.9 < F' < 103.2$.

STANDARDIZATION OF STATISTICS IN STATE INSTITUTIONS*

By Horatio M. Pollock, *New York State Hospital Commission*

The past half century has witnessed an enormous expansion of state institutions for the care of dependents, defectives, and delinquents. Prior to the Civil War there were in the United States but 24 state prisons or reformatories, 29 state asylums for the insane, and 44 state institutions for other classes. The last were principally schools for the blind and the deaf. The humanitarian movement inaugurated in the early forties by Dorothea L. Dix received new impetus upon the recovery of the states from the ravages of the war, and new state institutions rapidly sprang into being. State boards for the supervision of the several classes of dependents were created, and the policy of state care gradually became established. There are now about 600 state institutions in the United States, housing over 400,000 patients, inmates, or pupils, and representing an investment of over $400,000,000. The

*Read at the Eighty-second Annual Meeting of the American Statistical Association, Atlantic City, New Jersey, December, 1920.